Getting ready for indium gallium arsenide high-mobility channels

Mike Cooke reports on the VLSI Symposium, highlighting the development of compound semiconductor channels in field-effect transistors on silicon for CMOS.

Researchers across the world are readying the implementation of indium gallium arsenide (InGaAs) and other III-V compound semiconductors as high-mobility channel materials in field-effect transistors (FETs) on silicon (Si) for mainstream complementary metal-oxide-semiconductor (CMOS) electronics applications. The latest Symposia on VLSI Technology and Circuits in Kyoto, Japan in June featured a number of presentations from leading companies and university research groups directed towards this end.

In addition, Intel is proposing gallium nitride (GaN) for mobile applications such as voltage regulators or radio-frequency power amplifiers, which require low power consumption and low-voltage operation. Away from III-V semiconductors, much interest has been attracted around two-dimensional semiconductors such as graphene and, more recently, transition-metal dichalcogenides (TMDs) such as molybdenum disulfide. Massachusetts Institute of Technology (MIT) presented a number of developments, including a design flow, with possible future application of molybdenum disulfide in flat-panel display manufacturing.

InGaAs-OI

IBM has been working on two approaches to integrating InGaAs-on-insulator (InGaAs-OI) — confined epitaxial lateral overgrowth (CELO) and direct wafer bonding (DWB) — to achieve high-mobility channel devices on silicon.

The first demonstration of the CELO concept was presented by IBM Research's Zürich Laboratory in conjunction with the Swiss Federal Laboratories for Materials Science and Technology (EMPA) Electron Microscopy Center [L. Czornomaz et al, session 13-3]. Gate-first (GF) self-aligned fin-FETs were produced with electrical performance comparable to state-of-the-art InGaAs MOSFETs on silicon, according to the researchers.

The CELO process (Figure 1) starts by defining a seed area in a thermal oxide. The InGaAs is grown in a cavity. The cavity constrains the geometry, thickness and morphology of the InGaAs regions. By contrast, traditional techniques to reduce defects — e.g. epitaxial layer overgrowth or aspect-ratio trapping (ART) — are free in the vertical direction, and thickness and surface smoothness are determined post-growth by lithography or chemical mechanical polishing (CMP). The CELO process filters out defects by the abrupt change in growth direction from vertical to lateral, as constrained by the cavity. The researchers believe their technique avoids the main problems of alternative methods of integrating InGaAs into CMOS in terms of limited wafer size, high cost, roughness, or background doping.

The cap of the cavity was removed to access the InGaAs for device fabrication. Also the InGaAs material was removed from the seed region to electrically isolate the resulting devices from the underlying silicon substrate. At the same time, the 25nm-thick fins for the fin-FETs were etched. After this, a GF process flow was adopted. The gate stack consisted of aluminium oxide/hafnium dioxide (Al_2O_3/HfO_2) bilayer and tungsten (W) electrode. Devices were produced with and without raised source-drain structures.

For 150nm gate-length fin-FETs without raised source–drain structures, the subthreshold swing (SS) was 130mV/decade and the on/off ratio was 10^4, limited by gate leakage. The on-current (I_{on}) was improved to 0.4mA/μm for 80μm wide devices with raised source–drain contacts. Unfortunately, the off-current (I_{off}) was also increased through more gate leakage.

The researchers comment: "A transconductance benchmark against GF, replacement-gate (RMG) and gate-last (GL) InGaAs MOSFETs integrated on silicon reveals that the performance of CELO-integrated GF devices exceeds the one of similarly sized ART-integrated GF MOSFETs and compares to ART-integrated RMG devices."

The IBM Research Zürich Laboratory also led the work on DWB, which created ultra-thin-body InGaAs-OI structures [V Djara et al, session 13-5]. Also involved in this presentation were Université Grenoble Alpes and CEA, LETI MINATEC Campus, in France, and IBM T.J.Watson Research Center and IQE in the USA.

The researchers claim this as the first demonstration of ultra-thin-body (50nm), low-defectivity 200mm InGaAs-OI fabricated by DWB.

The InGaAs was first grown on a 200mm silicon donor substrate using molecular beam epitaxy (MBE) with a series of metamorphic layers bridging the lattice mis-match between the materials. The insulator consisted of an Al_2O_3 atomic layer deposited on the InGaAs and target silicon substrate. After bonding of the Al_2O_3 surfaces, the donor wafer and metamorphic buffer layers were removed with wet and dry etching, leaving a 250nm layer of InGaAs. The channel layer was further thinned to 50nm by CMP.

The researchers fabricated planar and fin-FET devices using GF and replacement metal gate (RMG) processes. The capacitive equivalent thickness (CET) of the gate insulator was as low as 1.2nm. Capacitance–voltage measurements suggested that the interface trap density (D_{it}) of the RMG process ($1.5x10^{12}/cm^2$-eV) was almost a factor of three lower than for GF ($4x10^{12}/cm^2$-eV). The researchers comment: "These results are among the best D_{it} versus CET reported for high-k/InGaAs."

In general, the RMG performed better than GF in terms of I_{on}, subthreshold swing and drain-induced barrier lowering (DIBL). A hydrogen/argon anneal improved SS and DIBL for both device types, although there was a slight degradation of I_{on} for GF transistors due to a shift in threshold voltage.

Noting that the RMG device type has a 'very competitive' trade-off between transconductance and SS, the researchers claim a record I_{on} against other InGaAs devices integrated on silicon of 118μA/μm at 0.5V operation with the I_{off} pegged at 100nA/μm. The team also claims that there was no evidence of short-channel effects down to 50nm gate length.

The researchers have also worked on scalability of the devices to small overall dimensions, rather than just the channel length, achieving a contact-to-contact pitch of 120nm. With a contact length of 70nm, the drain current was 261μA/μm at 0.5V drain and 1V gate. The contact resistance for 90nm length was 83Ω-μm.

Tunneling and quantum wells

Pennsylvania State University (PSU) presented results claiming record performance for complimentary all III-V heterojunction vertical tunnel FETs (HVTFETs) and a demonstration of InAs single and dual quantum well (QW) heterostructure fin-FETs (FF).

The HVTFET work was carried out with University of California Santa Barbara (UCSB) and the US National Institute of Standards and Technology (NIST) [R. Pandey et al, session 15-3].

Tunnel devices can achieve steep switching with low SS values. However, I_{on} and I_{on}/I_{off} ratios can be hit by the device structure. Performance improvement has been achieved by using antimonide (Sb) and arsenide (As) materials.

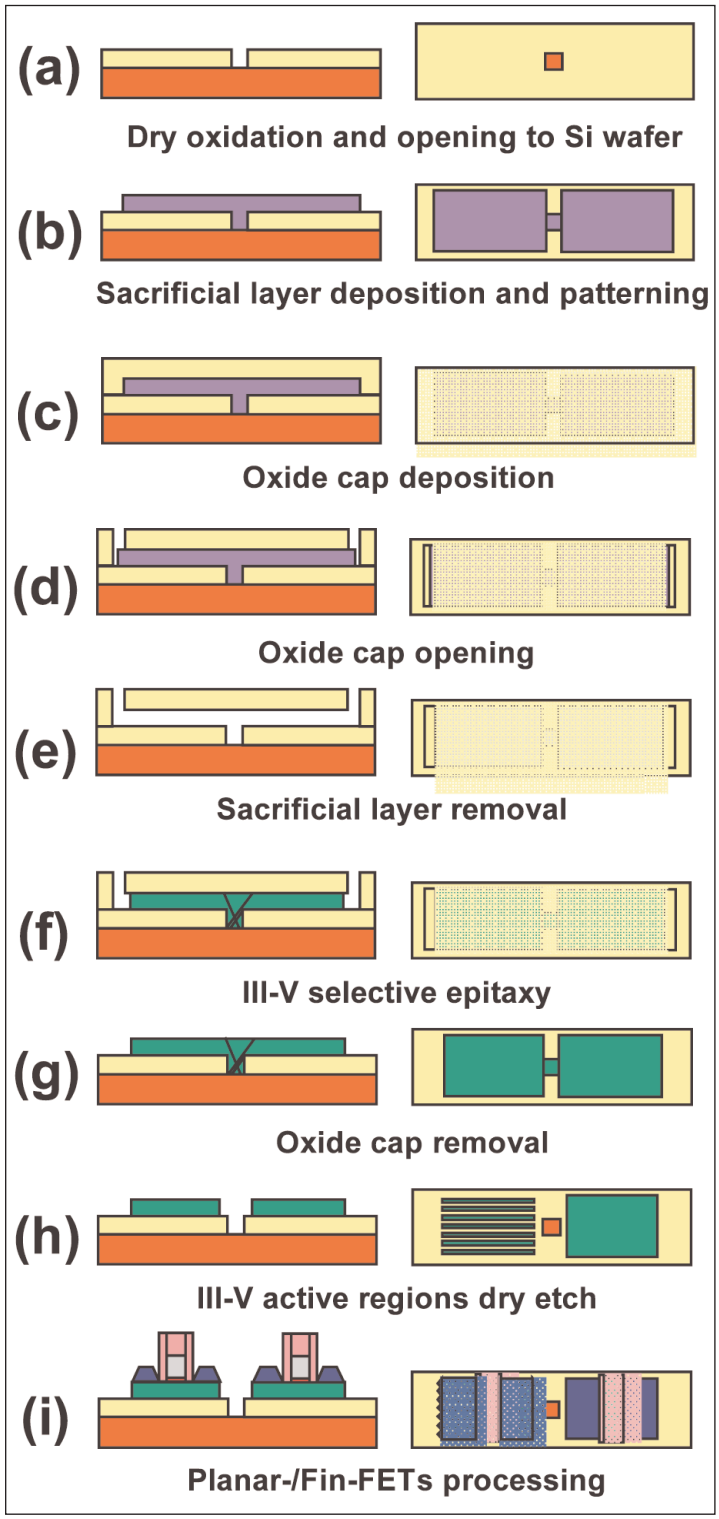


Figure 1. Process flow for integration of lattice-mismatched materials on insulator on Si by CELO method. Process only requires standard Si substrates and typical CMOS modules.

The researchers comment: "To implement energy-efficient complementary logic, both NTFETs and PTFETs need to be realized preferably in the same material system. Here, for the first time, we demonstrate complementary TFETs with high I_{on}, high I_{on}/I_{off} in

arsenide–antimonide material sharing the same metamorphic buffer layer."

The heterostructure material for the devices (Figure 2) was grown on a common metamorphic buffer on indium phosphide. However, the gate insulators for the n- and p-type TFETs differed, being zirconium dioxide and HfO_2, respectively, to match the different material contents of the channels.

The processing of the high-k dielectrics for the gate insulators had to be optimized. The PTFET needed a 150ºC hydrogen plasma clean to remove native oxide before applying 3.5nm HfO_2 to give a CET of 1.2nm and reduced mid-gap D_{it}. The NTFET insulator consisted of 4nm ZrO_2, giving a CET of 1.1nm and low mid-gap D_{it}.

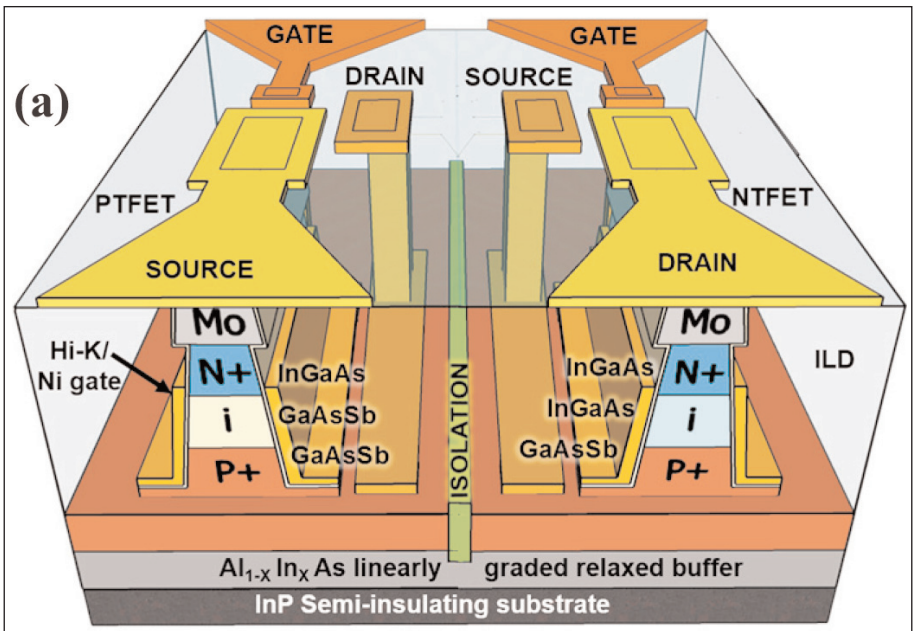
The PTFETs achieved an I_{on} of 30μA/μm and I_{on}/I_{off} ratio of 10^5 with drain bias of 0.5V. The corresponding results for the NTFET were 275μA/μm and $3x10^5$. The SS values were above the 60mV/decade limit value (kT/q) for planar devices at room temperature.

Pulsed operation reduced the swing by suppressing the response of slow mid-gap traps. The researchers report: "We achieve SS=55mV/decade for NTFET and SS=115mV/decade for PTFET at room temperature. The high I_{on} with sub-kT/q SS demonstration for NTFET and high I_{on} with improved SS demonstration in case of PTFET, is a direct consequence of engineering high-quality scaled gate dielectrics and tunnel barriers in the As–Sb system."

Comparing with TFETs produced with silicon, silicon-on-insulator or silicon germanium, the researchers note that they achieve higher I_{on}.

For the InAs QW fin-FETs, PSU teamed up with Samsung Electronics Co Ltd of the Republic of Korea [Arun V. Thathachary, session 15-4]. The aim of the research was to produce n-channel devices with superior performance compared with InGaAs fin-FETs.

InGaAs has higher mobility at higher indium concentration — hence InAs would theoretically offer the best high-mobility channels. However, strain considerations limit InAs to layers less than 5nm



(b)

PTFET	NTFET
400nm $In_{0.7}Ga_{0.3}As$ N+ (Si-5X10^{19} cm^{-3})	200nm $In_{0.65}Ga_{0.35}As$ N+ (Si-10^{18} cm^{-3})
200nm $GaAs_{0.35}Sb_{0.65}$ Intrinsic	150nm $In_{0.65}Ga_{0.35}As$ Intrinsic
500nm $GaAs_{0.35}Sb_{0.65}$ P+ (C-10^{19} cm^{-3})	10nm $GaAs_{0.4}Sb_{0.6}$ P++ (C-10^{20} cm^{-3})
100nm $Al_{0.3}In_{0.7}As$ Buffer	300nm $GaAs_{0.4}Sb_{0.6}$ P+ (C-5X10^{19} cm^{-3})
$Al_{1-X}In_XAs$ linear grade relaxed buffer	$Al_{1-X}In_XAs$ linear grade relaxed buffer
InP Substrate	InP Substrate

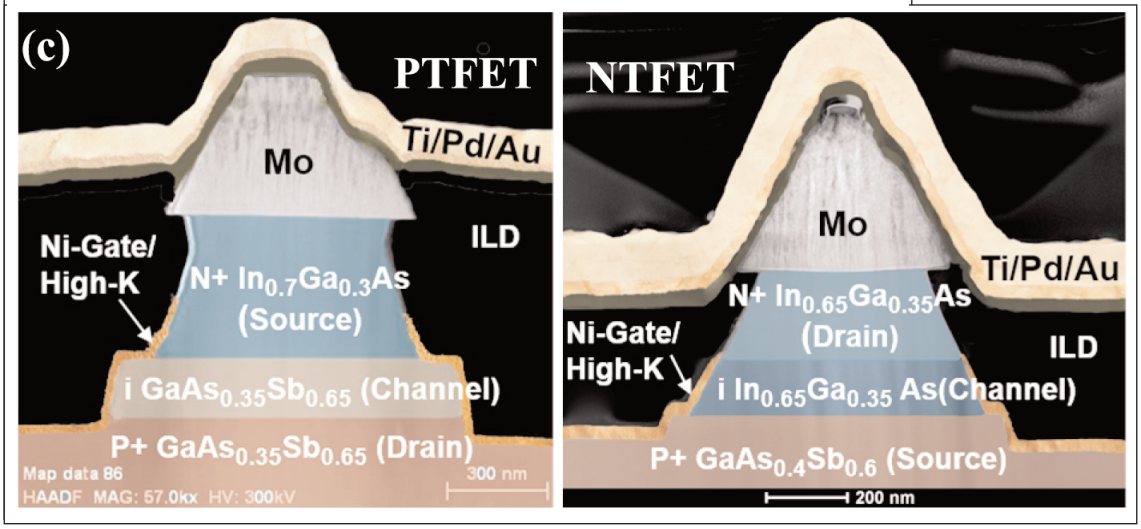


Figure 2. (a) Schematic of complimentary PTFET and NTFET on common metamorphic buffer technology; (b) starting heterostructures, and (c) cross-sectional TEMs.

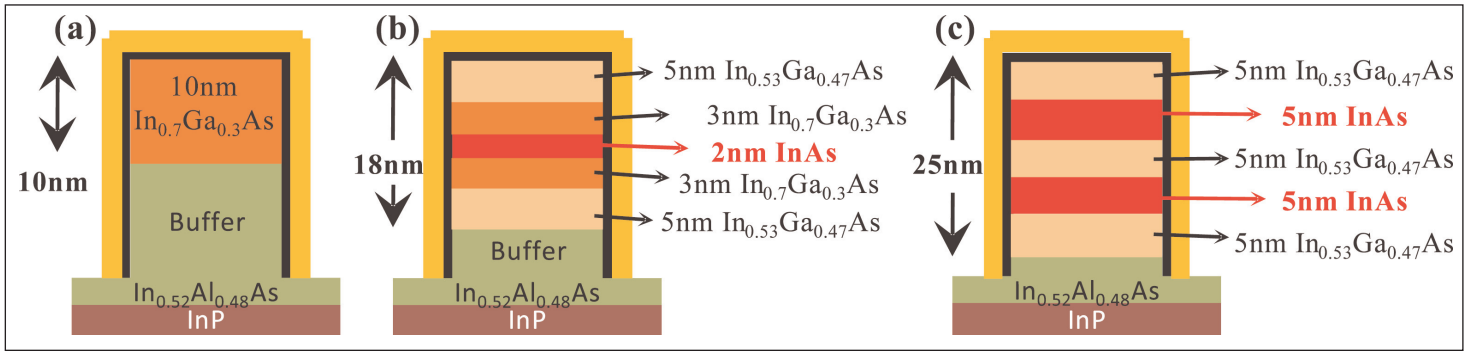


Figure 3. Cross-sectional schematic of (a) InGaAs QW, (b) InAs QW, (c) InAs dual QW fin-FETs.

thick when grown on indium phosphide (InP) substrates.

PSU and Samsung produced single and double InAs QWs by molecular beam epitaxy on InP for fabrication into fin-FETs (Figure 3). The fins were produced using a 'new side-wall image transfer (SIT) process'. The resulting fin pitch was 105nm. The high-k gate insulation was produced using alternate pulses of nitrogen plasma and trimethyl-aluminium to passivate the fin surface, followed by atomic layer deposition (ALD) of 3.25nm HfO_2. The gate metal was nickel.

The subthreshold swing for 2μm gate length was 87mV/decade for single QW fin-FETs and 94mV/decade for dual QWs — these values are described as 'excellent' by the researchers. The effective mobilities and carrier densities for 5μm devices were 3531cm^2/V-sec, 2x10^{11}/cm^2, and 3950cm^2/V-sec, 3.2x10^{11}/cm^2, respectively.

Projections to 26nm gate length suggest that the dual QW fin-FET would achieve an I_{on} 478μA/μm at an I_{off} of 100nA/μm and 0.5V drain bias, compared with 417μA/μm for a silicon fin-FET.

TSMC reported on $In_{0.53}Ga_{0.47}As$-channel metal-oxide-semiconductor (MOS) FETs fabricated on 300mm silicon [M. L. Huang et al, session 15-2]. The researchers claim Hall electron mobility values comparable to those achieved on InP. Devices with 150nm gate length achieved subthreshold swing of ~95 mV/decade, I_{on}/I_{off} ratio ~10^5, DIBL ~51 mV/V at 0.5V drain. "The extracted high field-effect mobility (μ_{EF} = 1837cm^2/V-s with equivalent oxide thickness (EOT) ~0.9nm) is among the highest values reported for surface-channel $In_{0.53}Ga_{0.47}As$ MOSFETs," the researchers write.

Mobile gallium nitride

Intel Corp researchers have been exploring "for the first time" the potential of gallium nitride transistors for low-power-consumption mobile system-on-chip (SoC) electronics [H.W. Then et al, session 15-1]. The researchers see their GaN MOS-HEMTs as being competitive against transistors used in voltage regulator (VR) and power amplifier (PA) applications.

Up to now, much of the focus in GaN transistor development has been toward high-voltage power switching and radio-frequency amplification. Intel's devices are normally-off 'enhancement-mode' metal-oxide-semiconductor high-electron-mobility transistors (MOS-HEMTs) rather than the normally-on depletion-mode Schottky gate HEMTs mostly used in the high-voltage applications.

The Intel transistors (Figure 4) used a barrier of aluminium indium nitride ($Al_{0.83}In_{0.17}N$) polarization layer on aluminium nitride (AlN) spacer to create two-dimensional electron gas (2DEG) source–drain access regions in the underlying GaN. The carrier density in the 2DEG was 2x10^{13}/cm^2 with mobility of 1200cm^2/V-s and sheet resistance of 250Ω/square.

The AlInN was etched away in the gate region. The gate insulation consisted of a high-k composite of 4.5nm HfO_2 and 1.5nm Al_2O_3 deposited on the AlN spacer, which was used to avoid mobility degradation in the gate region. The equivalent oxide thickness of the high-k stack was 23Å. The source–drain contacts

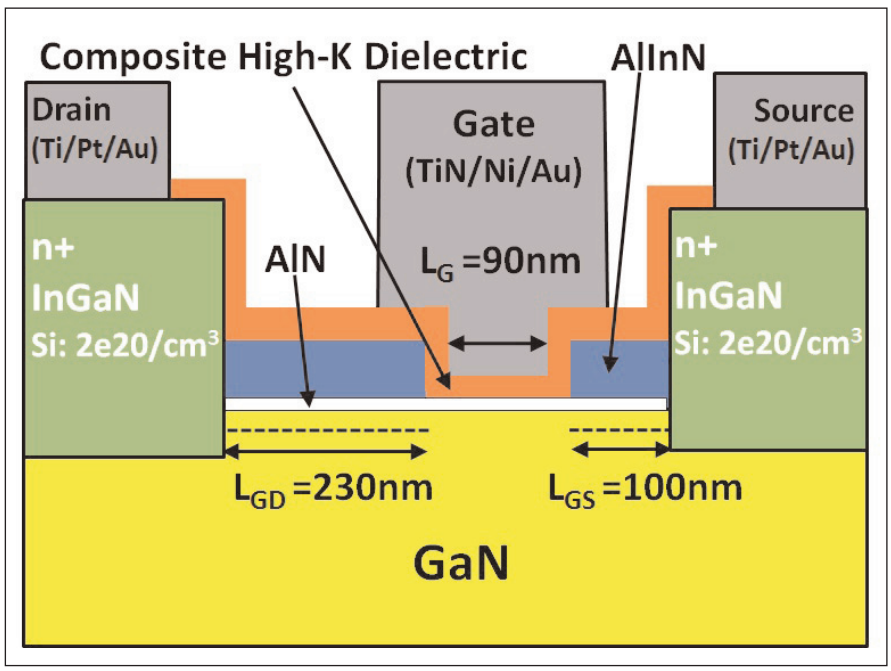


Figure 4. Schematic of enhancement-mode high-k GaN MOS-HEMT.

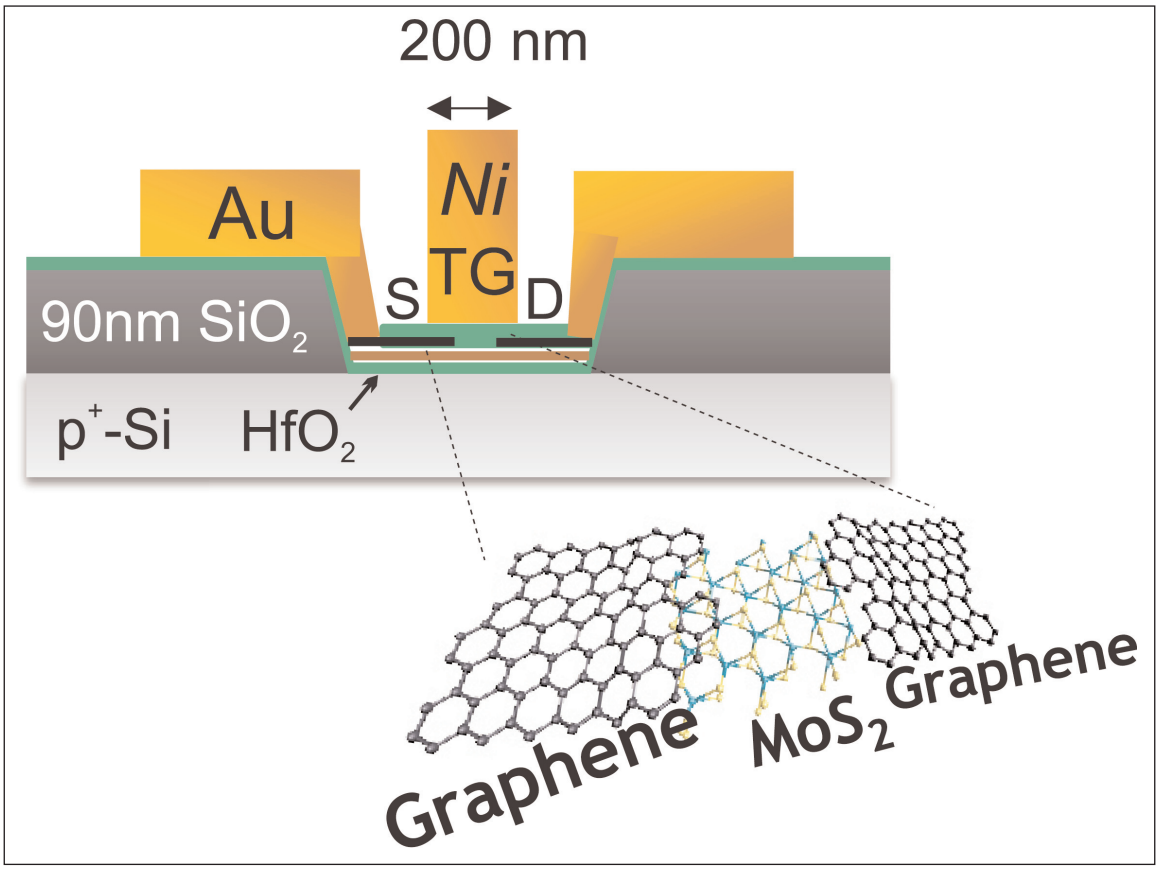


Figure 5. Schematic cross-section of short-channel single- and dual-gate-MoS_2 FETs with graphene source–drain contacts.

consisted of re-grown n^+-$In_{0.1}Ga_{0.9}N$.

Intel claims the best enhancement-mode GaN MOS-HEMT characteristics ever reported. These include a low I_{off} of 70nA/μm with 3.5V drain bias and 0V gate for a 90nm gate-length device. The drain-induced barrier lowering was 45mV/V. The maximum drain current was 1.4mA/μm with a knee voltage less than 1V. The on-resistance was 490Ω-μm.

The off-state breakdown (1μA/μm) for a 90nm-gate device and 230nm gate-drain distance was 8V. Comparing this with industry-standard silicon VR transistors with the same breakdown, the researchers comment that their GaN device has 3.6x lower on-resistance.

In frequency measurements, a 90nm-gate GaN transistor achieved a cut-off (f_T) of 100GHz and maximum oscillation (f_{MAX}) of 150GHz, "which exceed the requirements (f_T, f_{MAX}>20GHz) for 2G/3G/4G cellular applications at mobile SoC voltages," according to the Intel team.

For RF amplification, the researchers measured an 'excellent' power-added efficiency (PAE) of 80% with 0.55W/mm power density at typical mobile SoC voltages (3.5V drain bias).

The researchers comment: "Our GaN data show >10% better PAE at matched RF P_{out} or >50% higher RF P_{out} at matched PAE than industry-standard GaAs RF PA transistors at mobile SoC voltages."

Molybdenum disulfide

Massachusetts Institute of Technology (MIT) presented two papers on molybdenum disulfide (MoS_2) devices. The first detailed work with imec and KULeuven in Belgium on single- and double-gate FETs down to 15nm channel length [A.Nourbakhsh et al, session 3-4]. The second paper described efforts towards fabrication and modeling of large-scale flexible electronics circuits [Lili Yu et al, session 10-4].

MoS_2 is one of a range of transition metal (e.g. molybdenum or tungsten) dichalcogenides (e.g. sulfur or selenium) or 'TMDCs' that crystallize in layered structures. These layers can be separated rather like graphene from graphite using Scotch tape, allowing structures with two-dimensional electron transport to be created. As with graphene, there are alternative growth processes such as chemical vapor deposition (CVD) that could lead to mass production.

The MIT/imec/KULeuven single-/double-gate devices (Figure 5) were based on CVD monolayer or Scotch tape mechanical exfoliation of 4-layer MoS_2. The MoS_2 was transferred onto 10nm HfO_2 dielectric on p^+-Si on silicon dioxide (SiO_2) substrates. This gave a back-gate structure. The source–drain electrodes consisted of patterned monolayer graphene.

A top gate stack was fabricated by depositing 1nm Al_2O_3 in an air oxidation process, followed by atomic layer deposition of 10nm HfO_2. The gate electrode was 50nm nickel.

The MoS_2 was found to have n-type conduction in capacitance–voltage measurements. The FET operation was therefore accumulation-mode. A monolayer FET with 1μm channel length achieved an I_{on}/I_{off} ratio of 10^7 and a minimum SS of 75mV/decade. Four-layer devices had smaller I_{on}/I_{off} ratios and high SS (105mV/dec). Use of a top gate reduced the SS for the 4-layer MOSFET to 66mV/decade.

The team did not produce monolayer double-gate devices because the high-k dielectric deposition process

'drastically' shifts the threshold voltage and I_{off}. The 4-layer devices were less affected by the gate-stack fabrication.

Monolayer devices with channel lengths less than 30nm were affected by short-channel effects, giving high values of I_{off}.

A 4-layer MoS_2 double-gate FET with 15nm channel length at 0.5V drain bias achieved ~10^6 I_{on}/I_{off} ratio, 90mV/decade SS, and an I_{on} of 50μA/μm.

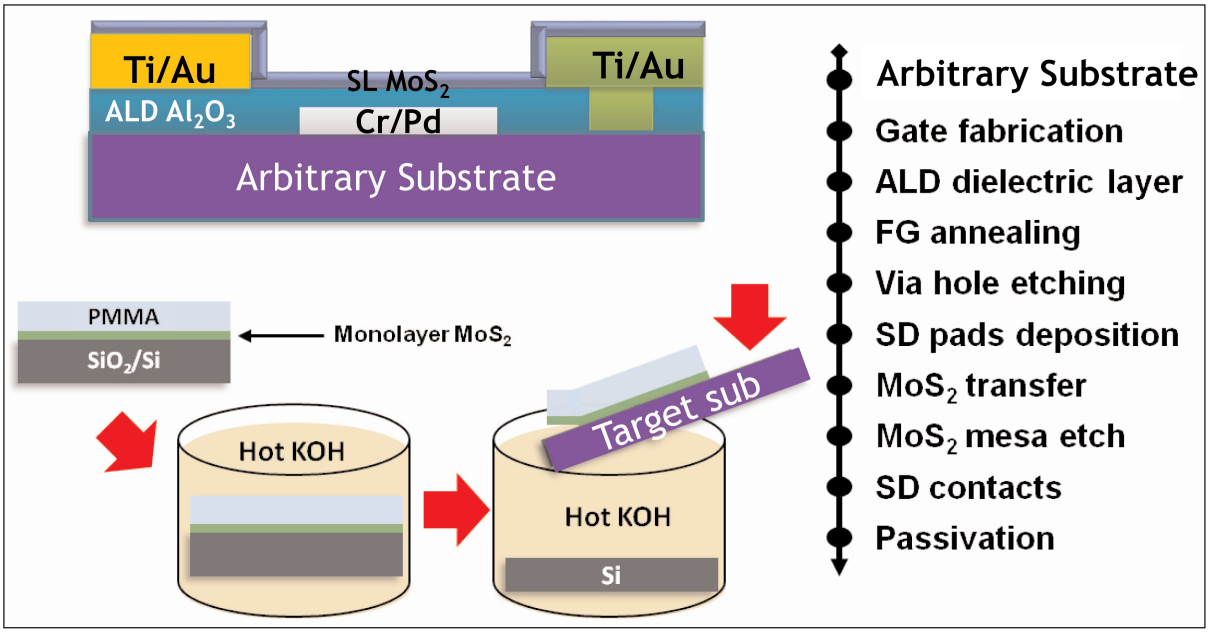


Figure 6. Fabrication technology for large-scale MoS_2 electronics.

The researchers comment: "This transistor has the shortest operating channel length of any MoS_2 transistor to date. The device performance indicates further scaling to sub $L_{S/D}$ = 10nm is possible."

The second MIT presentation of large-scale MoS_2 circuits used CVD growth and a GF fabrication technique to overcome the problems with producing the gate stack. The researchers point to flat-panel drive circuits as a possible application based on higher carrier mobility, compared with amorphous silicon or organic semiconductors, and a 1.8eV bandgap, which should give higher I_{on}/I_{off} ratios.

The team summarizes the challenges: "Despite its promising characteristics, applications up to date have been limited to single or a few devices scale system. The challenges in large-scale system design with a newly introduced material mainly remain in three issues: material growth and transfer, device integration, and circuit simulations."

The CVD process used sulfur and molybdenum trioxide (MoO_3) as precursors in a quartz tube furnace. A seed promoter of perylene-3,4,9,10-tetracarboxylic acid tetrapotassium salt (PTAS) was used to start the process. The substrate was SiO_2 on silicon.

The resulting monolayer MoS_2 demonstrated good uniformity and coverage approaching 100%, according to the researchers. The domain size was around 30μm on average.

Devices (Figure 6) were fabricated by first depositing the gate electrode on a SiO_2 layer on a silicon substrate. The gate electrode was covered with 20nm of Al_2O_3 by atom layer deposition. After annealing, a via hole was etched to give access to the gate electrode.

The MoS_2 monolayer was coated with poly(methyl methacrylate) (PMMA) thermoplastic before being released by soaking in 85ºC potassium hydroxide solution to etch away the SiO_2. The MoS_2 was then transferred to the device substrate and annealed to clean away the polymer residue. Mesa isolation was achieved with oxygen plasma etch to form the MoS_2 channel. Ohmic contacts were formed with 5nm titanium and 90nm of gold. Contact pads were created with 90nm gold.

The researchers say that they used the GF method to avoid fixed charge and trapped states inside the dielectric or at the interface with the MoS_2, which degrade mobility and shift threshold voltages. The problem with GL techniques is the low temperature or seeding processes needed to avoid damaging the underlying structure. "Using the gate-first process instead, the critical components are fabricated before the MoS_2 transfer step, achieving high gate dielectric and interface quality and the potential for low effective oxide thickness scaling," the team comments.

GF transistors had 10x larger I_{on} and 100x smaller I_{off} than GL devices, along with a higher threshold voltage – 0.54V (enhancement-mode, 0.12V standard deviation) rather than –4.20V (depletion-mode, 1.75V standard deviation) for GF. The subthreshold swing was 115mV/decade for the GF transistors, compared with 300mV/decade for a GL device.

The researchers have also developed a computer-aided design flow along with circuit simulation based on a Verilog-A compact model for the transistors. The flow was used to create a switched capacitor DC–DC converter from two MoS_2 transistors and a charge transfer capacitor. The simulation predicted the performance of the circuit well with switching frequencies up to 100kHz. ■

Author: Mike Cooke is a freelance technology journalist who has worked in the semiconductor and advanced technology sectors since 1997.